

# Машинное обучение

## ФКН ВШЭ

### Теоретическое домашнее задание №2

**Задача 1.** Найдите константу  $C$ , решающую следующую задачу ( $0 < \tau < 1$  фиксировано):

$$\sum_{i=1}^{\ell} \rho_{\tau}(y_i - C) \rightarrow \min_C,$$
$$\rho_{\tau}(x) = \begin{cases} \tau x, & x > 0, \\ (\tau - 1)x, & x \leq 0. \end{cases}$$

**Задача 2.** Покажите, что если в задаче регрессии  $p(y_i|x_i, w) = \frac{\alpha}{2} \exp(-\alpha|y_i - w^T x_i|)$  (распределение Лапласа,  $\alpha$  фиксировано), то метод максимального правдоподобия эквивалентен оптимизации MAE для модели линейной регрессии.

**Задача 3.** Представим, что в некоторой задаче мы можем разбить признаки на  $k$  непересекающихся групп (например, такие группы возникают при использовании one-hot кодирования — по одной группе бинарных признаков на каждый категориальный признак). Кроме того, мы хотим в модели линейной регрессии задать свой коэффициент  $L_2$ -регуляризации для каждой группы. Какому априорному распределению на веса это будет соответствовать?

**Задача 4.** Убедитесь, что вы знаете ответы на следующие вопросы:

- Почему  $L_1$ -регуляризация производит отбор признаков?
- Почему коэффициент регуляризации нельзя подбирать по обучающей выборке?
- Что такое кросс-валидация, чем она лучше использования отложенной выборки?
- Почему категориальные признаки нельзя закодировать натуральными числами? Что такое one-hot encoding?
- Для чего нужно масштабировать матрицу объекты-признаки перед обучением моделей машинного обучения?
- Почему MSE чувствительно к выбросам?

- 
- Какие методы можно применять для оптимизации MAE?
  - Что такое Huber Loss? В чем его преимущества по сравнению с MAE и MSE?
  - Почему квантильная регрессия так называется?
  - Какова общая постановка задачи линейного программирования?